Key

Chapter 10 Review

1. What are explanatory and response variables?

2. What is the graph that relates 2 quantitative variables?

   Scatterplot

3. How do we describe a scatterplot?

   form
   direction
   strength

4. What is the formula for the LSR line?

   $\hat{y} = a + bx$

5. What do each of the symbols/variables in the LSR line formula stand for?

   y-intercept   slope

6. What is r? What does it measure?

   correlation – how linear a relationship is (and what direction)

7. What is r^2? What does it measure (how is it interpreted)?

   coefficient of determination. __% of the Δy is due to Δx (or LSR line).

8. What is a residual? How is it calculated?

   error   obs – predicted y-values

9. On the calculator:
   a. How do we find the LSR line?

   STAT → CALC → 8: LinReg a+bx

   b. How do we find the residuals?

   $_L$RESID (after LinReg is done)

   c. How do we create a scatterplot?

   Statplot – x vs. y   then Zoom #9

   d. How do we create a residual plot?

   statplot – x vs. $_L$RESID

   e. How do we find r and r^2?

   LinReg → it comes up automatically

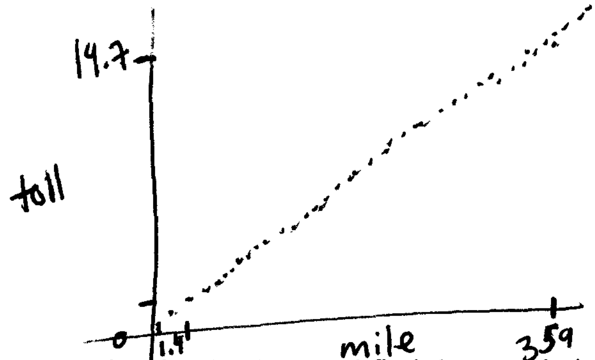
PIKETOLL, 83g
(MILE, TOLL)

AP Statistics
Section 2.3 – Linear Regression (Turnpike Tolls)

Tolls for the Pennsylvania Turnpike and the mileages corresponding to them (measured from west to east) are contained in the lists MILE and TOLL in the group file PIKETOLL.

1. Create a scatterplot of the data. Which variable would likely be the explanatory variable?

describe
- positive
- linear
- strong



Exp l = mile
resp = toll

2. Find the LSR line for the data, and find the correlation coefficient. Does the line model the data well? Explain.

yes. High r and r^2

$\hat{y} = -0.1157 + 0.0401x$

$r = 0.999$

3. Interpret the slope of the line. What is the significance of the y-intercept?

slope = for every mile, \$0.04 (4¢) ↑

int = 0 miles ≈ − \$0.12

4. What proportion of the variability in turnpike tolls is explained by the regression line with mileage?

$r^2 = 99.83\%$

5. Use the regression equation to predict the toll for a person who needs to drive 150 miles on the turnpike.

$\hat{y} = -0.1157 + 0.0401(150) = \$5.91$

6. By how much does the regression equation predict the toll to rise for each addition mile that you drive on the turnpike?

slope = $\frac{0.04}{1}$ ~ 4¢

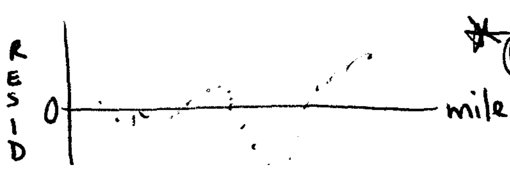
7. About how many miles do you have to drive in order for the toll to increase by one dollar?

$\$1 = 0.04x$ x = about 25 miles

8. Do there appear to be any influential points or outliers? Explain.

no

9. resid plot



*another model is better.

What is the form for the **sample** regression line *(the LSR line)*?

$\hat{y} = a + bx + e_i$

What is the form for the **population** regression line?

$y = \alpha + \beta x + \varepsilon_i$

Statistic	Parameter	What is it?
a	α	intercept
b	β	slope
$\hat{y}$	y	resp. variable
e_i	ε_i	residuals

What is the sum of the residuals? 0

What are the mean and the standard deviation for the residuals of an LSR line?

$\mu = 0$ $\sigma = s =$ depends on data

How do we estimate the standard deviation of the residuals? (symbol and formula)

$$s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Testing the Significance on β

What are we testing? slope of population regression line

Hypotheses: Ho: $\beta = 0$

Ha: $\beta \;\substack{>\\<\\\neq}\; 0$

Test Statistic: $t = \dfrac{b - \overset{=0}{\beta}}{SE_b}$

P-value: $P(t \;\substack{>\\<}\; \text{test stat} \mid df = n-2) = tcdf(LB, UB, df)$

df = $n-2$

Conclusion: We reject/fail to reject.....

We have suff. evidence that... the slope of the pop. regr. line is $\underline{\substack{>\\<\\\neq}}$ 0.

Thus... as $\underline{x}$ increases, $\underline{y}$ $\underline{\text{incr.}}$

decr.
changes
doesn't change

$>$
$<$
$\neq$
$=$

Conf. Interval: $b \pm t^* \cdot SE_b$

t^* from table

Conclusion:

We are ___% conf. that the slope of the pop. regr. line btw. $\underline{x}$ and $\underline{y}$ is btw $\underline{a}$ + $\underline{b}$ units/units.

Assumptions: 1) 2 indep SRS

2) true relationship is linear

Examples:

2 ways to test the slope:

1. **From computer output**

Variable	Coeff		Std. Dev.
Constant	-5.76	a	1.234
Year	0.348	b	0.0445

s= 3.1923 R-sq= 91.22% R-sq(adj)= 87.67%

s = std. dev. of resid

R-sq $= r^2$

n= 40

$H_0: \beta = 0$

$H_a: \beta \neq 0$

$t = \frac{b}{SE_b} = \frac{0.348}{0.0445} = \boxed{7.820}$

$2 \cdot P(t > 7.820 \mid df = 38) =$

$= 2 \cdot tcdf(7.820, E99, 38)$

$= \boxed{1.9667 \times 10^{-9}}$

- reject
- $\beta \neq 0$
- as x increases, y changes

2. **From the actual data**

a. Put data... in L_1 and L_2

b. On calculator... Lin Reg T test

t =

p =

df =

a =

b =

s = std. dev. of resid $\neq SE_b$

r^2 =

r =

- If asked to find SE_b...

$t = \frac{b}{SE_b}$ ... then solve

over →

Airfare

$H_0: \beta = 0$

$H_a: \beta > 0$

$\hat{y} = 83.267 + 0.1173x$

$r = 0.794$

$r^2 = 0.632$

$t = \frac{b}{SE_b} = 4.1442$

$P(t > 4.1442 \mid df = 10) = 9.993 \times 10^{-4}$

- reject
- $\beta > 0$
- as distance increases, so does airfare